

# Weekly Report

Aug 20<sup>th</sup>, 2018

## Done:

1. 论文套路完成到 visual design 一半. 量比较大, 下周继续. 具体内容往下翻
2. 追着马老师讨论了三次, 目标是: 在给定多数据源的时空数据下, 对于不同需求, 用一种自动机制生成 visual report. 我们不会设计过于复杂的可视化方法, 重点是这个自动机制, 如何满足需求, 如何安排空间, 时、空那个是主要维度? 下周继续. 思考总结这些问题

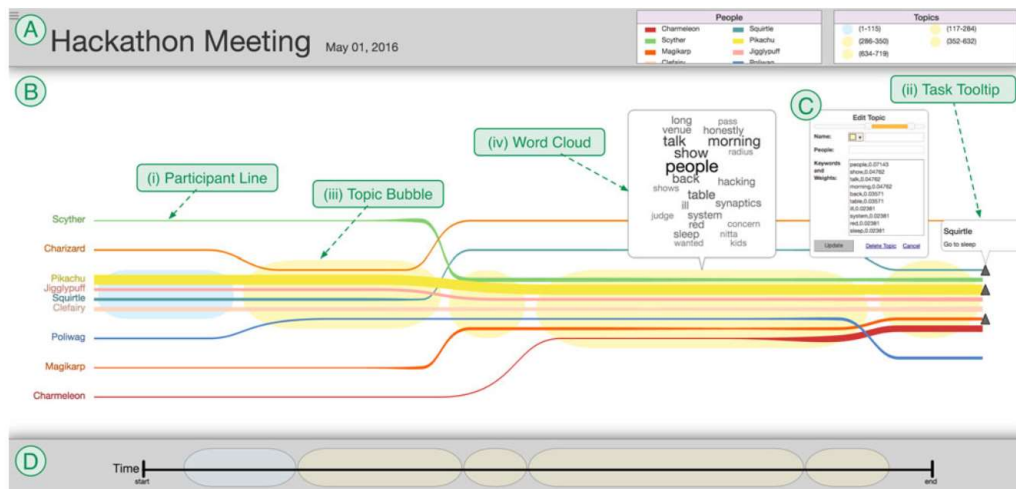
## To Do:

1. 如上面所述的未完项目, 以及研究生素质 ppt.

## Paper Reading:

VIS11 *Exploratory analysis of time-series with ChronoLens* 面向对象可以是单个单变量时间序列, 两个单变量时间序列(同一个序列不同时间戳), 也可以是多变量时间序列(合一或者 stacked). 对于时间序列的各种操作; 对于时间序列间的各种操作, 对一个序列的操作可以 apply 到另一个上 (横向); 可以自己 pipe 一个操作的 hierarchy(纵向). 感觉这个系统很简单但很实用, 写的也很理论化.

TVCG18 *MeetingVis: Visual Narratives to Assist in Recalling Meeting Context and Content* 将会议记录分析需求巧妙的映射到了 storyline 的可视化上, 可以看会议文字细节, 讨论主题, 成员参与度, 时间分布, 问题解决等, 也是十分有用.



## Design Study 套路整理

Junhua Lu

注 1: 蓝色下划线表示可选表示这是部分文章有的结构.

注 2: 文章列表

[SellTrend] SellTrend: Inter-Attribute Visual Analysis of Temporal Transaction Data

(TVCG09); **VIS09 best paper honorable mention**

[Log] Combing the Communication Hairball: Visualizing Parallel Execution Traces using Logical Time (TVCG14, VIS14)

[VariantView] Variant View: Visualizing Sequence Variants in their Gene Context (TVCG13, VIS13)

[TenniVis] TenniVis: Visualization for Tennis Match Analysis (TVCG14, VIS14)

[SoccerStories] SoccerStories: A Kick-off for Visual Soccer Analysis (TVCG13, VIS13) 写法过于特殊

[Weaver] Visually Comparing Weather Features in Forecasts (TVCG16, VIS15)

[Poemage] Poemage: Visualizing the Sonic Topology of a Poem (TVCG16, VIS15)

[SignalLens] SignalLens: Focus + Context Applied to Electronic Time Series (TVCG10, VIS10)

[SnapShot] SnapShot: Visualization to Propel Ice Hockey Analytics (TVCG12, VIS12)

[Vials] Vials: Visualizing Alternative Splicing of Genes (TVCG16, VIS15)

## Abstract

Vials 先抛出专业定义与技术, 要干什么难点在哪. 然后引出他们设计. 他们设计能帮助用户完成什么? 最后再说我们用 **case study** 验证系统的有效性(或者价值).

SnapShot 虽然本文是冰球分析, 但是 **abstract** 站在体育分析的角度上讲现状, 然后先 **identify** 体育分析时候可视化可以切入的地方, 然后讲系统, 然后讲特定的一个设计(毕竟是 **design study**). 最后提通过 **user study** 来自各方面的反馈, 证明有效性.

开门见山: 提出了什么, 与什么数据有关、要解决什么问题, 现有技术采用了哪些, 有什么不便? [SoccerStories]; 我们的系统采用了什么技术; 用了什么方法证明系统的有用性, 获得了什么评价 [, SoccerStories] [, SellTrend]

背景启发式: 伴随着某种技术 A 的发展, 相依的技术 B 遇到了越来越大、各种各样的挑战. 这些挑战使得某种分析方法 C 非常重要. 这种方法 C 往往是用到可视化技术的, 但是现有的可视化 D 又不太好, 会有哪些问题. **As an alternative, we present a new approach based on XXX.** 然后简述这种方法; 最后用什么证明了系统的有用性, 这个证明结果是怎样(好的反馈、容易得到 **insight** 等) [, Log, SnapShot, TenniVis, SignalLens]. 可以不用这么 ABCDE 这么深的 **hierarchy**, 可以浅一点; 当然这仍然与你本文贡献点有关, 与你这个项目是否是领域首创、或者比较容易 **generalize** 有关. [VariantView, Vials, Weaver, Poemage]

## Introduction

理论上来说, **introduction** 是 **abstract** 的一个扩充版本; 会更加详尽的描述 **abstract** 中提到的各项. 对于比较专业的领域, 他已经要开始解释一些名词了, 比如 [Vials] 这种讲基因相关的, 但是仍然不是细讲, 是点到一些名词为止, 可能会用图片来辅助解释, 能让读者顺利读下去掌握到问题的精髓. 因为正如上面所说 **design study** 都是专门领域设计的, 需要有专门的大段 **Background** 来介绍.

**一般套路:** 都是先提出大背景. 这个背景一般是领域相关的; 然后提出常有技术以及现有技术的不足; 然后提到可视化, 如果现有对策已经有可视化的则还要提他们的不足 | 如果没有, 而可视化在这方面很有价值, 则直接继续写可视化在这个领域为什么有优势. 然后就是本文提出的方案, 其特点是什么. 如何做的 **evaluation**. 最后概括其贡献. 文章其余部分介绍.

**变体:** 介绍大背景、小背景后, 列举 一些 challenges, 这些 challenges motivate 我们去设计了 XXX. [Weaver, Poemage]

**非常规套路 1:** [SoccerStories]用很短的篇幅写了大背景和现有技术, 然后便提到了一个概念. 这个概念既是足球分析中的一个概念, 也是本文设计的一个基础, 所以一开始就提出了. 他并没有归纳现有技术的问题. 但是, 在 **Related Work** 中他有仔细分析各种现有技术的特点与不足.

**非常规套路 2:** 其实是属于一般套路的分支: 在 **introduction** 里先提出 goals 的文章 [TenniVis, Log]: 一般的文章是直接在遇到的挑战之后直接写我们做的设计与系统的介绍; 但上述文章增加了 goals list ([TenniVis]是一个宏观的 goal list, 对整个专家交流、设计、反馈流程; [Log]是传统的 design goals), 来表明我们通过观察现有的技术后的一个思考的框架, 然后和别的文章一样写设计与系统的介绍 (这时候的介绍注意也要扣紧上面的 goals). 这样, 逻辑上其实更清晰, 读下来会更顺利; 唯一可能的问题是会让 **introduction** 篇幅过长.

## Design Process

仅在 [VariantView, WeaVER, Poemage, SellTrend] 出现了, 前三者出现的位置在 **Background** 之前; 其作者分别包含 Tamara Munzner(第一)和 Miriah Meyer(第二、三). 这两位大师各自及合作了一些理论性的框架性的可视设计的文章, 其实就像指导思想一样很有意义:

- Bridging from Goals to Tasks with Design Study Analysis Reports
- Design study methodology: Reflections from the trenches and the stacks.
- Exploratory Sequential Data Analysis: Foundations
- Design Activity Framework for Visualization Design

其实他们肯定要一方面自己引用文章, 另一方面为了证明自己的框架有用所以必须要体现一下, 更何况这本来就是 design study 的文章.

本章充当了一个 meta analysis 的作用, 但可能最后也流于形式(我们平时见的、写的往往都是一小段带过; [SellTrend]虽然写的比较详细但其实也有这样的意思)、或者说本来就在行文之中隐式的体现了(文章的写作顺序算是对设计总体过程的一个体现), 所以很多文章没有写. **一般套路:** 我们与哪些专家进行合作, 这些专家分别来自于哪些领域的详细介绍. 我们的 design process 是采用了哪种框架? [VariantView, WeaVER]这两篇文章都用了上面第二种中的 nine-stage framework. 然后只要按照这个框架的顺序叙述即可. 可以再加上时间分配等信息. [Poemage]则更像是在讲一个故事, 怎样一步步与诗人沟通, 让他们排除对可视化或者计算机方法处理这种人文的东西的不屑一顾, 然后到他们被吸引过来, 一起讨论, 到确定研究方向等等. 这些确实是一个真正的 design study 应当具有的步骤; 目的是真正体现你这个东西是经过不断沟通交流, 拥有可信的专家参与与意见的作品.

## Background (Data and Tasks)

*注: (个人想法)为什么列为第二节? 因为大多数文章都是位于此. 可以这么想: 虽然常规 related work 都是第二节的, 但考虑到 design study 多为 domain problems, 直接上 related work 仍然会出现一些专业术语, 如果 introduction 里没有讲, 那只能让人云里雾里, 无法继续读下去了.*

**Background** 显得尤为重要, 因为数据多是来自于 domain, 很难懂必须要专门独立出来一节; 目的是先要讲清楚各种数据的专业术语、他们数据之间各种关系; 其次最好是把数据变成计算机科学领域的说法(data abstraction, data model), 这样也方便去构建后面的 tasks. 不同

文章的 tasks 也可能会以多种形式出现(比如说成 workflow 或者用户感兴趣), 但都紧扣主题、紧扣用户需求的.

**一般套路:** [VariantView]直接将 data 和 tasks 放在了同一个 section, 各自一个 subsection. 强调是从和用户的交流、反馈中提取的. Data 按照一个自顶向下、从大到小的顺序描述(俄罗斯套娃); Tasks 先依据现有分析师的做法按照不同任务先分类; 再凝练、细分出小的一句话描述的任务(Q1-Q13), 作为后面可视化设计要遵循的东西. [SellTrend]也放在了一起, 但是由于相对简单, 叙述比较扁平. 类似的, [Vials]将这个套路变成了两个 sections; 在 data 这部分做了一个小的总结(Data Model, 有些文章称为 data abstraction[WeaVER]). 推荐最后抽象好以后的数据, 若类型繁多, 用表格的形式呈现出来[VariantView, Poemage]

**拓展:** [VariantView][Poemage]作为两篇强调的写了 design process 章节的文章, 在描述这个 background 的时候, 也充分体现了专家的意见、与专家的交流. 这也让文章更有理有据令人信服.

[Log]一文将 background 和 related work 放一起了, 个人不太推荐.

**变体:** 前面提到的 Introduction 中已经提出 goals 的 TenniVis, 他在叙述背景知识时候顺便把(1) 对于比赛分析十分重要的信息 (2) requirement (加粗)也一起写了. 同样的, 这样读起来很顺畅, 能让读者对不管是领域本身还是 requirement 有更深刻的印象. 这也是一个不错的写法.

例子: [TenniVis] **2 Background and requirement analysis:** As with many sports, there is a psychological component to tennis that affects players differently. Some players ease off in their intensity after getting a lead just to see their lead slipping away (known as **choking**). Other players are able to step up and recover after being down in a match (known as **rallying back**). **Choking and rallying back should be easy to spot in visualizations because they may identify player patterns that need to be modified.**

**例外:** [SoccerStories]的没有对应上述 background 的, Problem Description 中写背景的小部分写的过于简单; 后面更多描述了宏观上的分析需求. 至于可视化层面的需求或者任务在 design 那边 visual design 讲到, 下面会提及.

## Related work

**一般套路:** 都是按照一定的分类方式, 对每个类的现有系统进行概括. 写作时一定要市场点名关键点, 表明你的系统也做了或者做的更好.

常用的划分的方式:

- 有的按照**数据类型**划分[SellTrend]: 如[SellTrend]按 time series、categorical、faceted metadata 划分.这个我们比较常用, 但是在这些文章里并不是如此. 可能是要考虑到这个 domain specific 的原因, 一般不会做一个十分 generalizable 的设计或者系统, 所以不按照这个来.
- 有的按照分析**数据哪些方面(分析需求)**来划分, 如[SoccerStories]中, 分析足球时候需要从哪些角度来看? [TenniVis]也类似
- 有的按照**可视化技术**或者**交互技术**来划分[SignalLens]: 3.1 Context-based Visualization Systems 3.2 Lens Framework 3.3 Time Series Visualizations and Analytics; [WeaVER]分成天气可视化与不确定性可视化两类.

只要和自己所做的东西密切相关, 言之成理即可, 上述分类也可混合或者嵌套[Vials].

## Visual design[有些会直接用设计的名称作为 section 标题]

这一部分样式繁多.

常见的是总分结构来写

总分结构 TenniVis 4.1 4.2 不一样 后面 view 应该是根据需求的吧?

非常潇洒的写法: [VariantView]

[SellTrend]

**例外: 非常规套路** [SoccerStories]它是一个两步设计法, 先设计一些小视图(faceted views for soccer), 再用某些可视化方法与交互串起来(SoccerStories, 真正一个系统); 给人的感觉就是集成创新. 这样写的原因是, 其在 background 中的描述过于宏观了, 没有深入到数据层面. 因此直接到小时图一节才提出一个数据概要的表格, 并用图片展示.

[Log]一文也是先用一个 section 讲述他们的核心设计 Logical Time, 其背后的数理理论、可视设计又是如何满足这些理论的. 然后再讲系统的各个部分; 同样的, 这时候也可以采用先概述各个 view 功能, 概述用户交互, 再细节写每一个 view 各自的设计及附带的交互.

[Vials]有个 design principle, Our design is guided by principles partially motivated by the domain problem and partially based on visualization theory and experience.